

特許明細書作成のための論文、特許明細書からの情報抽出手法

大屋 由香里[†] 新森 昭宏[†] 谷川 英和[‡]

難波 英嗣^{††} 奥村 学^{‡‡}

インテック・ウェブ・アンド・ゲノム・インフォマティクス株式会社[†] IRD国際特許事務所[‡]

広島市立大学^{††} 東京工業大学^{‡‡}

1. はじめに

近年、企業や大学では特許の出願が盛んであり、一般の技術者や大学職員が特許明細書の作成に関わる機会が増えた。

特許明細書は一種の法的文書であり、独特の記述形式を持っている。そのため、弁理士や知的財産権担当者などの専門家以外の人にとって馴染みにくく、記述の仕方に戸惑うことが多い。また、出願作業は非定型的な作業であり[1]、生産性が良いとはいえなかった。

そこで著者らは、特許明細書作成の足がかりとして、既存の特許明細書や発明者の作成した論文を利用することが有効であると考えた。論文には、特許出願したい発明が簡潔にまとめられ、かつ、その説明が詳細に行われている。技術者にとって特許明細書よりも論文の方が書きやすく、弁理士にとっても、論文は発明の理解に役立ち、作成にかかるコストを削減することができる。

また、発明の品質の維持は、もっぱら個人のスキルに頼っている[1]。既存の特許明細書は、特許明細書を作成するためのノウハウが詰まっており、技術者にとっては特許明細書の作成の参考となり、弁理士にとっては効率的かつ品質の高い特許明細書作成のための材料となる。

本稿では、論文から、発明の詳細な説明に関する情報を抽出する手法、および既存の特許明細書から、特許明細書の記述のための表現や情報を抽出する手法について述べる。

2. 論文からの情報抽出

論文からの情報抽出手法の検討に当たり、まず、特許明細書と論文の類似性について調査した。調査対象は、特許法第30条第1項の適用申請がある特許明細書とそれに対応する論文である。特許法第30条第1項とは、新規性喪失の例外に関するものであり、特許出願する前に発明の内容を第三者に発表してしまっても、特許法第30条第3項で指定された学会等で、かつ出願前6ヶ月以内のことであれば新規性喪失には当たらない、とするものである。つまり、特許法第30条第1項が適用されている特許には、その技術について述べた論文が必ず存在しているということであり、その論文こそが本稿において、特許明細書作成の足がかりとなると考えるものである。これらの比較結果を表1に示す。

表1 特許明細書と論文の記述の類似性

特許明細書項目	論文中の記述
発明の属する技術分野	序論の自身の研究に関する記述。
従来技術	序論の他者の研究に関する記述。
発明が解決しようとする課題	序論の他者の研究の問題点に関する記述。
発明の実施の形態	研究の詳細、評価について述べたセクション内の記述。
発明の効果	—

よって、論文から上記に該当する部分が特許明細書作成において有効な情報である。これらを抽出するための手がかりとして、文中に含まれる語や表現を用いる。調査に用いた論文を教師データとして、手がかり表現を検討した。主なものを表2に示す。

表2 論文からの情報抽出の手がかり表現

論文中の記述	手がかり表現
自身の研究に関する記述	「著者」、「本稿」等の自身を指す表現

An information extraction method for generating patent specifications from research papers and patent specifications

[†] Yukari Oya and Akihiro Shinmori INTEC Web and Genome Informatics Corporation

[‡] Hidekazu Tanigawa IRD PATENT OFFICE

^{††} Hidetsugu Nanba Hiroshima City University

^{‡‡} Manabu Okumura Tokyo Institute of Technology

論文中の記述	手がかり表現
他者の研究に関する記述	「……されている」等の第三者的な表現
他者の研究の問題点に関する記述	「問題」、「できない」等の見解を表す表現
研究の詳細・評価についての記述	「実験」、「検証」等の実施に関する表現

3. 特許明細書からの情報抽出

冒頭でも述べたように、特許明細書の記述には独特の表現方法が用いられる。また、発明の構成要素の説明が十分に行われている必要がある。これが不十分であると、審査において拒絶の原因となり、さらに、不用意な記述が特許範囲を狭めることにもつながる。

これらの問題には、特許電子図書館[†]等で入手できる特許明細書が参考となる。発明の説明が十分に行われるよう、その記述のための定型表現や説明手順、また、特許権の権利範囲を狭めることが無いよう、発明の実施方法を網羅するための発明の部品の実現方法を抽出する。

特許明細書内に、頻繁に出現する文パターンを定型表現として抽出する。複数の文において、文末（述語）から文頭に向かって、係り受けを考慮した順に文節単位で、語が一致するかを調べ、一致する箇所がある場合、その語の並びを一つの文パターンとする。

発明の部品の実現方法については、文中に含まれる語や表現を手がかりに抽出する。主なものを以下に示す（「○○」が部品、「…」が実現方法である）。

- …等の○○手段
- …で○○する
- ○○手段は…でも良い

なお、これらの表現は特許明細書の項目「発明の実施の形態」に現れる。よって、検索範囲はこの項目とした。

4. 検証

4.1. 論文からの情報抽出

2で述べた手法を実装し、検証した。

使用した論文は、手がかり表現の検討の教師データとは別に収集した、特許法第30条第1項が適用されている特許明細書の適用理由となったもの10件である（主に情報処理学会研究報告）。結果は、項目により差があり、「従来の技術」が再現率約80%、精度約70%で最も良く、「発明が解決しようとする課題」が再現率、精

度ともに約50%で最も悪かった。「発明が解決しようとする課題」に該当する論文の記述には、これと比べて特徴的な表現が無く、抽出が難しい。なお、全体的に精度が低かったが、元々の抽出量が少ないため、特許明細書作成者が読んで確認しても負担になるようなものではなく、あまり大きな問題ではないと考えている。ただ、「発明の実施の形態」について、現在のセクションレベルでの抽出では粒度が大きすぎるので、今後はさらに絞り込むための条件を検討したい。

4.2. 特許明細書からの情報抽出

3で述べた手法を実装し、検証した。

定型表現の抽出の検証に使用した特許明細書は、弊社出願のもの（ソフトウェア）を弁理士別に集めたものである（1弁理士あたり5件前後で4弁理士について）。抽出された表現は、一部の語句が欠けているものもあるがほぼ期待通りに抽出できた。ただ、出現頻度に依存するため、特に定型表現とは呼べない記述も抽出されてしまう。特に「発明の実施の形態」において定型表現とはしないものが抽出結果の約40%を占め、ノイズが非常に多い結果となってしまった。

発明の部品の実現方法の抽出の検証については、抽出したい部品名を含む特許明細書10件を用いて行った。結果、部品名が「○○手段」と記述されている場合の実現方法はほぼ期待通りに抽出することができたが、「○○する」と記述されている場合はノイズが有効データより多い結果となってしまった。抽出条件について再検討する必要がある。

5. おわりに

本稿では、論文および既存の特許明細書から特許明細書の作成に有益な情報を抽出する手法について述べた。抽出結果には、ノイズが多いが有益な情報も高い再現率で含まれている。今後は抽出条件をさらに検討し、精度を高めたい。

謝辞

本研究開発は、独立行政法人情報処理推進機構（IPA）2006年度次世代ソフトウェア開発事業に関わる委託業務に基づくものです。

参考文献

- [1] 谷川英和, 河本欣士, 特許工学入門, 中央経済社, 2003.
- [2] 谷川英和, 田中克己, 3種類の特許部品データベースに基づく特許明細書自動生成エンジンの構築, 情報処理学会論文誌 Vol. 46 No. 30, 2005.

[†]<http://www.ipdl.ncipi.go.jp/homepg.ipdl>